A CONFIDENCE LEVEL ALGORITHM FOR THE DETERMINATION OF ABSOLUTE CONFIGURATION USING VIBRATIONAL CIRCULAR DICHROISM OR RAMAN OPTICAL ACTIVITY.

Elke Debie[1,2] , Ewoud De Gussem[3,4,5] , Rina K. Dukor[2,4], Wouter Herrebout[4,5], Laurence A. Nafie[1,2,4], Patrick Bultinck[3,4,]*

[1] Department of Chemistry, Syracuse University, 1-014 Center for Science and Technology, Syracuse, New York 13244-4100, USA

[2] BioTools, Inc., 17546 Bee Line Highway, Jupiter, Florida 33458, USA

[3] Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281- S3, 9000 Gent, Belgium. Patrick.Bultinck@UGent.be

[4]The European Centre for Chirality, www.chiralitycentre.eu

5 Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerpen, Belgium

Abstract

Spectral comparison is an important part of the assignment of the absolute configuration (AC) by vibrational circular dichroism (VCD), or equally by Raman Optical Activity (ROA). In order to avoid bias caused by personal interpretation, numerical methods have been developed to compare measured and calculated spectra. Using a neighbourhood similarity measure, the agreement between a computed and measured VCD or ROA spectrum is expressed numerically to introduce a novel confidence level measure. This allows users of Vibrational Optical Activity (VOA) techniques (VCD and ROA) to assess the reliability of their assignment of the AC of a compound. To that end, a database of successful AC determinations is compiled along with neighbourhood similarity values between the experimental spectrum and computed spectra for both enantiomers. For any new AC determination, the neighbourhood similarities between the experimental spectrum and the computed spectra for both enantiomers are projected on the database allowing an interpretation of the reliability of their assignment.

Keywords

Chirality, configuration determination, IR, CD, VCD

I. Introduction

The active pharmaceutical ingredient (API) of many drugs and drug candidates is chiral. Although until recently drugs with chiral APIs were largely commercialized as racemates, more recently single-enantiomer APIs are taking a commanding priority in the market. This extraordinary interest in chiral, single-enantiomer molecules as APIs calls for very reliable methods for the assignment of their absolute configuration (AC). As the absolute configuration of a molecule may have very significant consequences for the activity of the molecule as a drug, international regulatory agencies have become increasingly strict and require both enantiomers of a chiral drug candidate to be tested separately for their therapeutic and adverse effects before the actual drug can be launched[1]. Over the past decades vibrational circular dichroism (VCD) and Raman optical activity (ROA), jointly called Vibrational Optical Activity (VOA) methods, in combination with *ab initio* calculations, have become reliable methods for the assignment of the AC of chiral molecules in the solution phase[2,3].

Unfortunately, it is not possible to directly deduce structural information from an experimental VCD spectrum. Therefore, theoretical spectra for a postulated absolute configuration, computed using reliable quantum chemical algorithms developed over the past decades[4,5], are used to establish the link between the experimental spectrum and the AC. Based on the agreement between both spectra, meaning the location, sign and intensity of the bands, the AC can be determined. However, measured spectra are inherently different from results of a quantum mechanical calculation. Difficulties in assigning the AC can arise for more complex molecules exhibiting high conformational flexibility or tendency towards self aggregation[6,7], form complexes with solvents[8-10] or show other effects that are normally not

or insufficiently well described by the theoretical methods. On a more technical level, the many approximations taken in the computational approach, such as the use of a finite basis set and, at the Density Functional Theory level, the functional, also have an impact. A trained eye is often used to assess the degree of agreement between computed and experimental spectra while taking into account the possible complications described above. “Human eye” comparison, however, may be very subjective and can be biased by personal interpretation especially when the agreement is only of intermediate quality. In order to avoid such bias and further improve the reliability of VCD, additional numerical comparisons should be performed. The classical but very time consuming method is to individually link IR fundamentals in the theoretical and experimental spectrum. Then one derives rotational strengths from the experimental VCD bands and determines the correlation between measured and calculated rotational strengths with emphasis on the agreement in sign. A review of such successful assignments can be found in Stephens[11].

Up until now very few studies have been published that discuss criteria to assess whether a reliable assignment of the AC has been made. In reference 12, Minick explains the quality assessment used at GlaxoSmithKline: “the key to confident predictions at GSK is the value of the [...] coefficients of correlation between the intensities of 10 to 15 corresponding bands in the calculated and measured spectra. Our assignments are considered reliable if r^2 is at least 90%”. Such an approach is rather tedious because individual bands need to be cross-linked between theory and experiment and because choosing a reduced number of bands in establishing the absolute configuration may introduce some arbitrariness. In search of a more direct and immediate quantification of the agreement, a neighbourhood similarity (NS) measure has previously been introduced for VCD[13-16]. This similarity measure, based on the overlap between measured and calculated spectra, takes into account the neighbourhood in the direct proximity of the bands. Despite earlier fruitful applications of such NS measures[13-16],

no test of such measures has been performed over a wider set of molecules for which the AC is known from other approaches such as explicitly correlating VCD bands between theory and experiment or other experimental techniques such as X-ray diffraction. Moreover, no confidence level has yet been established for VCD assignments. This paper addresses both issues. First we apply the NS and related measures, described below, to a larger set of molecules and secondly we establish a confidence level for the assignment of the AC of a new molecule. Although in this paper only VCD will be discussed, the proposed algorithm is also considered valuable for the comparison of ROA spectra.

II. Theoretical Background

A. *Neighbourhood similarity (NS)*

The similarity measure, used in this study for the quantification of the agreement between the calculated, $f(\nu)$, and measured, $g(\nu)$, IR or VCD spectra, is a fairly straightforward adaptation of a cosine based similarity measure as will be demonstrated in detail below and as was also used by Kuppens *et al.* [13-16], following up on work by De Gelder *et al.* [17] for comparing powder diffraction spectra. Instead of such a generalized cosine in a multidimensional vector space, one can also use the arithmetic mean as normalizing term or, in fact, any other similarity measure, including the Tanimoto one as recently used by Shen *et al.* [18]. As the similarity measure used in the present paper differs somewhat from previous expressions based on the generalized cosine, we opt for a re-derivation and a detailed description of the algorithm.

We first establish what parameters a program for VCD spectrum similarity should contain to allow meaningful similarity measures:

1. It is well-known that computed absorption frequencies are overestimated. The

common approach to correct for this effect is to introduce a global scale factor. We therefore introduce a scaling factor σ that acts on the theoretical spectrum $f(\nu)$. The scale factor depends on the chosen theoretical method and lies for most cases between roughly 0.89 and 1.00[19]. The algorithm described in the present paper allows for optimization of this parameter in the way detailed below.

2. Although advances in spectrometers have reduced the uncertainty in the position of the baseline substantially, some degree of uncertainty may remain. We therefore allow a uniform shift of the baseline. The highest absolute VCD intensity in the experimental spectrum is sought and the shift is varied between $-x$ and $+x$ where x equals a user chosen percentage of this absolute value. Experience has shown that 10% is often a good choice. The baseline shift is then optimized in the way described below and applied over all frequencies in the experimental spectrum.
3. Even when a scale factor is introduced, one cannot account for all possible local shifts. Different techniques have been implemented to take such effects into account. First of all, the theoretical spectrum for every conformation of a molecule is a line spectrum and in order to make the global molecular theoretical spectrum mimic more closely the experimental spectrum, the collection of the theoretical conformational spectra is combined using Boltzmann weighting and each line in the resulting molecular line spectrum is broadened using a Lorentzian bandshape. In the next step, the global scale factor is applied. To account for small local shifts required to bring the theoretical and experimental spectrum closer to each other, instead of using a uniform scale factor, one could use different factors for different types of vibrations and thus perform a different scaling for each type of vibration in the spectrum[20]. This obviously requires having different scaling factors for different vibrations which again requires individual assignment of bands to correlate them to a specific type of vibration and establishing

some sort of database of scale factors. This is a rather tedious procedure. Therefore, in this work small local shifts are taken in to account in a different way. Instead of using a point by point comparison between the theoretical and experimental spectrum, we arrange that the experimental spectrum at every point bears to some degree information on the surrounding frequencies and their accompanying VCD intensities. This is done by replacing the original experimental spectrum at every frequency ν_0 by one where the intensity of the surrounding points is included. These points are weighted in such a way that their contribution becomes smaller as they are farther from ν_0 :

$$g(\nu_0) \leftarrow \int g(\nu)\, w(\nu - \nu_0)\, d\nu \qquad (1)$$

The weighting function $w(\nu - \nu_0)$ determines the extent to which the neighbourhood of ν_0 is taken in to account. In this study a triangular weighting function was chosen. This allows one to take into account the neighbourhood with a width of $-l$ and $+l$ around each point in the spectrum:

$$\begin{aligned} w(\nu - \nu_0) &= 1 - \frac{|\nu - \nu_0|}{l} \qquad & |\nu - \nu_0| \le l \\ w(\nu - \nu_0) &= 0 & |\nu - \nu_0| > l \end{aligned} \qquad (2)$$

The shape of the weighting function is chosen based on the fact that corresponding bands in the experimental and theoretical spectrum can be shifted locally from each other. The value of l has to be chosen carefully in order to avoid linking too distant bands. The default has been set to 20 cm^{-1} but can optionally be changed.

Taking in to account that we henceforth consider the theoretical spectrum to be the original computed spectrum with Lorentzian broadening and scaling and the experimental spectrum to be the original spectrum modified through (1) and with a shifted baseline, the

generalized cosine similarity expression is given as:

$$S_{fg} = \frac{\int_a^b f(\nu) g(\nu) d\nu}{\sqrt{\int_a^b f^2(\nu) d\nu \int_a^b g^2(\nu) d\nu}} \qquad (3)$$

Obviously this degree of similarity depends on the values for the parameters described above which can optionally also be optimized as will be described below. As is clear from equation (3), the similarity can be computed for any range of frequencies through choosing the lower and upper frequency (resp. a and b in (3)). If within this range a peak can be established to be an artefact, it should be replaced by zero intensity over the range of the artefact in both the theoretical and experimental spectrum so as not to bias the similarity in any undesirable way. The interesting aspect of equation (3) is that the similarity measure always lies within 0 and 1, provided that the intensity over all frequencies has the same sign. This is a consequence of the properties of vector spaces with an inner product as present in the numerator of (3). Equation (3) is readily applicable to IR spectra, henceforth denoted S_{fg}^{IR} , but needs to be adapted to be useful for VCD spectroscopy.

B. VCD neighbourhood similarity and enantiomeric similarity index.

Equation (3) is a normalized quantity with values within the interval [0,1] (or [0,100] as percent scale). Here 1 (or 100) corresponds to the comparison of identical spectra. Having an index that is for all molecules confined to the same range of values is very useful as it allows comparing results for different molecules. This would not be the case with for instance an Euclidean distance where no upper limit can be given.

For VCD spectra, computing the similarity between spectra is slightly more involved.

By the very nature of a differential spectroscopy, the intensity at a given frequency can have both signs, positive or negative. One could opt to simply use (3) to compute the similarity and not take the sign in to account in any special way. This has the drawback that the overlap between spectra in regions of equal sign between the theoretical and experimental spectra and regions of opposite sign can compensate in the integral in the numerator of (3), possibly leading to undesirable effects. In order to avoid this, we simply split the scaled theoretical and the experimental spectrum, into a positive and a negative spectrum. Following (3), a similarity measure is computed for the positive spectra only ($S_{fg}^{++,VCD}$) and one for the negative spectra only ($S_{fg}^{--,VCD}$). In order to have one single similarity measure, Σ_{fg}, a weighted mean of both similarities is computed as:

$$\Sigma_{fg} = \frac{\Phi^{++} S_{fg}^{++,VCD} + \Phi^{--} S_{fg}^{--,VCD}}{\Phi^{++} + \Phi^{--}} \qquad (4)$$

The weight Φ^{++} simply reflects the amount of VCD signal of specific sign in the theoretical and experimental spectrum, i.e. it is the sum of the surface of the positive theoretical and experimental spectra:

$$\Phi^{++} = \int_{f(\nu)>0} f(\nu)\,d\nu + \int_{g(\nu)>0} g(\nu)\,d\nu \qquad (5)$$

Φ^{--} is analogous but for the negative parts of the spectra. As $S_{fg}^{\pm\pm,VCD}$ lies always within the limits $[0,1]$ and $\Phi^{\pm\pm}$ is always a positive number, Σ_{fg} also lies in the range $[0,1]$ so that it can still be considered a similarity measure. In the same way as described above, one can also compute the similarity measure $\Sigma_{\bar{f}g}$ for the theoretical spectrum $\bar{f}(\nu)$ of the other enantiomer with respect to the experimental spectrum.

Kuppens *et al.* [13-16] suggested the difference between the neighbourhood similarities of the measured VCD spectra versus each of the corresponding calculated spectra of both

enantiomers as a criterion to assess the degree of success of a VCD assignment of the AC. This differential neighbourhood similarity measure, i.e. the enantiomeric similarity index, ESI, gives information about the discriminating power between the two enantiomers. In this work the absolute value of the ESI, henceforth denoted Δ is used throughout:

$$\Delta = \left| \Sigma_{fg} - \Sigma_{\bar{f}g} \right| = \left| ESI \right| \tag{6}$$

If the measurement and simulation of the spectra are reliable, the calculated VCD spectrum for one enantiomer should show good agreement with the measured VCD, while the spectrum of the opposite enantiomer should hardly show any agreement. This is the basis of the discrimination potential of VCD. Δ is limited to the interval [0,1]. High Δ values indicate that one of the computed enantiomeric spectra has a significantly better agreement with the observed spectrum $g(\nu)$ compared to the other. Low values indicate that $f(\nu)$ and $\overline{f}(\nu)$ have similar values of Σ and thus that assignment of the AC via VCD cannot be performed with high reliability. Note that the scale factor and the shift in the baseline are both determined as the values that give the largest Δ by developing both parameters in a simple grid. First the scale factor is optimized, followed by the optimization of the baseline shift.

Equation (3) shows that only the regions where the sign in the theoretical and experimental spectra agrees, contribute to the similarity. This means that regions in the spectra where the sign of the rotational strength in the theoretical and experimental spectrum differs do not play role. However, the introduction of Δ in fact introduces these regions in the following way. Starting from (6), we have:

$$\Delta = \left| \frac{\Phi_{fg}^{++} S_{fg}^{++,VCD} + \Phi_{fg}^{--} S_{fg}^{--,VCD}}{\Phi_{fg}^{++} + \Phi_{fg}^{--}} - \frac{\Phi_{fg}^{-+} S_{fg}^{-+,VCD} + \Phi_{fg}^{+-} S_{fg}^{+-,VCD}}{\Phi_{fg}^{-+} + \Phi_{fg}^{+-}} \right| \tag{7}$$

where we used the mirror image relationship between the VCD spectra of two enantiomers. Δ thus introduces the effect of having regions in the computed and experimental spectra that do not agree in sign.

Although numerical measures similar to the one introduced in this paper had been tested in several cases[13-16], their use must be validated in order to allow them to be used more universally. In this study, this is done by testing them on a much larger set of molecules for which the AC has been determined correctly. Therefore, Σ and Δ values have been computed for such a database. If these results reveal that they can be used reliably, a confidence level measure can be introduced (see below) by projecting the values of the similarity for the best fitting enantiomer, Σ^{max}, and the accompanying Δ of a new VCD experiment in the set of known values from the database. If the new assignment would fall outside the range of pairs of Σ and Δ for previous successful AC determinations, the new assignment may be unreliable.

III. Results

A. Test database for VCD similarity analysis

The AC of 84, mostly pharmaceutical compounds was previously assigned after thorough analysis of the agreement between measured and calculated IR and VCD spectra using the elaborate peak by peak comparison and regression between dipole and rotational strengths extracted from theory and experiment. Moreover, for several molecules, the assignment of the AC was confirmed by other techniques such as X-ray diffraction. For each of these assignments the similarity measures introduced above were computed as well as the value for Δ from equation (6). The window size, l, of the triangular weighting function in eq. (1) was set to 20cm^{-1} for all calculations. In general, this value has proven to give overall larger enantiomeric discrimination power. The factor used to scale the frequencies of the calculated spectra is varied numerically between 0.89 and 1.15 until the maximum absolute value of Δ has been found. In general only the range between 1000cm^{-1} and 1500cm^{-1} has

been considered to compute similarity measures because carbonyl-stretching vibrations, which occur in the range of 1600 to 1700 cm^{-1}, are often heavily influenced by the solvent and this functional group occurs ubiquitously in molecules submitted to VCD analysis. In order to include sufficient bands in the analysis, it is important that at least a 400 cm^{-1} span of frequencies be compared. As described above, the similarity measures may also be influenced by the position of the baseline of the observed VCD spectrum. In order to take into account a possible small offset from zero, the measured VCD spectrum was shifted up and down within a range of 10% of the maximal intensity, until a maximum Δ value was obtained.

For the database of 84 compounds the Δ value was optimized for every molecule and Σ computed for both enantiomers. The highest value of Σ, Σ^{max}, among both enantiomers is plotted as a function of the corresponding Δ value as shown in Figure 1. The blue markers represent instances of agreement between the enantiomer corresponding to Σ^{max} and the AC assignment based on thorough analysis of the measured and calculated VCD spectrum using elaborate band by band correspondence and comparison of the rotational strengths or an AC based on X-ray diffraction. By contrast, the red dots correspond to cases where the enantiomer corresponding to Σ^{max} does not agree with the result of the more thorough determination of the AC. By plotting a new molecule whose VCD spectrum has been measured and for which Σ^{max} and Δ have been computed, one can readily assess whether the assignment of the AC can be expected to be reliable or not. It is important to stress that the use of the present algorithm is not directed towards replacing the manual assignment of the AC by a numerical technique. The similarity analysis is aimed at providing the chemist an indication of whether the assignment made is at par in the level of confidence with the bulk of previous successful assignments. It also flags cases where the chemist might possibly want to reconsider the assignment made by e.g., checking the result for a different stereo-isomer, but also in that case a convincing manual assignment remains strictly required.

<Add figure 1 here>

Based on the plot in Figure 1, confidence levels can be set up for future new AC assignments. We numerically express the degree to which VCD can still be considered to have reliably assisted in the assignment of the absolute configuration by computing the ratio of the number of correct assignments with respect to the total number of attempted assignments around the position of the newly assigned AC of molecule a. Specifically, we choose the following scheme for a confidence level:

$$CL(a) = \frac{\sum_{i}^{N} e^{-\alpha d_{ia}^2} \delta_i \left(correct\right)}{\sum_{i}^{N} e^{-\alpha d_{ia}}} \qquad (8)$$

In this equation, d_{ia} is the Euclidean distance between a molecule i from the database and new molecule a in figure 1, N is the number of molecules in the database and $\delta_i \left(correct\right)$ is a logical construction that assigns a weight equal to one to a database compound i if the presently described algorithm was correct at giving a higher similarity for the enantiomer that corresponds to the one that was actually used in the VCD experiment. The exponent α has been chosen 0.01 based on the fact that tests revealed this gives a smooth transition from the area with mostly molecules where the algorithm indeed points out the same AC as the assigned one to the area with more wrong AC's. Wrong in this case is defined as a case where the algorithm did not successfully give the correct isomer the largest similarity. Inspection of figure 1 shows that for several molecules the highest similarity was found for the wrong AC. Many of those lie in an area where Δ is also low, meaning that there the difference in similarity of the computed spectrum with respect to experiment for the two enantiomers is low. This suggests that any user-made assignment of a molecule in that region, based on

correlation of bands, must be considered with extra care. There are 4 molecules that have relatively high Δ and Σ^{max} values but for which the algorithm led to the wrong conclusion although for the majority of points in that region of figure 1 the algorithm led to the correct enantiomer having the highest similarity. For several among these 4 points, we did find that the optimal scale factor, maximizing Δ, does have less common values such as exceeding 1.00. Still, such points are important as they delineate the area above which no wrong results occur.

The confidence level is a simple numerical value that basically expresses how far to the upper right a new assignment lies in figure 1 and supplements the conclusion that can be drawn based on the location of a new assignment in figure 1. The algorithm thus assists the chemist in answering the question to what extent a given assignment appears to be reliable, or in simpler terms: when a chemist has answered the question on the AC, the algorithm tells how much confidence can be put to this answer. If the degree of confidence is small, this may suggest trying a different assignment, which obviously also must be first checked by visual verification and analysis of the bands or may point out other difficulties like problems with basis set, incomplete representation of the conformational distribution etc. Figure 2 shows how *CL* with exponent 0.01 behaves when going from the lower left corner of the plot in figure 1 to the upper right corner. To show this in a simple 2D graph, a generalized coordinate of both coordinates in figure 1 has been made. This coordinate is denoted SQRT and is the geometric average of the two coordinates of every molecule in figure 1. The *CL* performs in the desired way by attaching the highest confidence only to the points where the generalized coordinate is high, i.e., it reflects how far to the upper right an assignment lies in figure 1.

<insert figure 2 here>

B. Similarity measures and confidence levels for (+)- 3R-methylcyclohexanone

As an example of the use of the data in figure 1 and the confidence level, CL, similarity measures and confidence levels are computed for the VCD based assignment of the AC of (+)-3*R*-methylcyclohexanone. A thorough analysis of the measured and calculated IR and VCD spectra of the molecule has already been discussed extensively in literature[21-23]. The aforementioned similarity measures of the compound were obtained after running Compare*VOA*[24], the program in which the here described algorithms have been implemented. In Figure 3, the measured IR and VCD spectra and the computed spectra (B3LYP/6-31G*) are shown. The intensities have been scaled to emphasize the agreement between both sets of spectra. The measured spectrum very clearly corresponds to the R configuration as is easily confirmed by visual inspection of the spectra and agrees with the mentioned previous assignments.

<Add figure 3 here>

The magnitude of the optimal scale factor (0.970) that maximizes Δ, lies within the range of scale factors used for harmonic spectra obtained with hybrid density functionals[19]. For $f(\nu)$ corresponding to the R configuration the IR spectra lead to $S_{fg}^{IR} = 60.4\%$. For VCD $\Sigma_{fg} = 67.55\%$ and $\Sigma_{\bar{f}g} = 3.94\%$ resulting in $\Delta = 63.61\%$. Based on the value of Σ_{fg} and Δ, it is concluded that the AC of the compound is *R* with confidence level equal to 98%. One of the more appealing aspects of computing the different measures for an AC assignment is that one can also plot the newly assigned molecule together with all molecules of the database as in Figure 1. It shows that the molecule, indicated by the green marker, indeed lies among all

previously correctly assigned AC's.

The computation of all required integrals to establish the confidence level can be performed in just a few seconds. Obviously, an automated approach to establishing the AC using the algorithm described here should not replace chemical correlation and expertise. The reliability of the new method clearly depends on the quality of the observed and calculated spectra and needs to be established by an experienced VCD user. In cases where experimental effects like intermolecular association occur but are not taken into account in the computations, one should not use this measure because theory and experiment are too far apart. However, in such cases the introduced measures tend to immediately classify the assignment as having a low confidence level and at the same time positions the molecule in the lower left corner of figure 1. We therefore suggest to always make sure that the following requirements are fulfilled: First experimental measurements should be performed according to best VCD practice. Second, a good level of theoretical calculation must be used. Then the numerical measures as presented here should be computed and can serve as a guide to possibly improve e.g., the calculation of the quantum chemical spectra or as an indication to look for other effects that may lower the degree of confidence. VCD expertise does remain a third requirement as some of the capabilities of CompareVOA must be used carefully, e.g., the possibility to exclude some bands from the similarity analysis. In all cases, a first assignment by an experienced VCD user is required after which CompareVOA can express the reliability of the assignment or possibly may suggest the user to consider a different stereochemistry or suggest that some aspects of the VCD analysis should be improved.

Comparing our similarity measures with the method developed by Shen *et al.* [18] some similarities but also some differences are observed. The algorithm of Shen *et al.* computes the spectrum similarities, SimIR and SimVCD, via a Tanimoto coefficient. The local shift of the bands is not taken into account through the use of triangular weighting functions, as done

here. In their approach, the scaled spectrum is divided into a number of bands and the frequency of each band is again shifted in search for the nearby maximal IR similarity. Finally the scaled and shifted bands are pieced together. A few bands, however, may need additional shifting, which is handled by a user-controlled shifting adjustment. This makes the method more laborious and possibly more open to bias. Secondly, the obtained VCD similarity value ranges from -1 (opposite enantiomer) to 1 (correct enantiomer) and it is suggested that in order to establish high confidence AC determination the associated absolute value of the similarity measure should be greater than 0.2. In our CompareVOA program, however, the study of a data set of 84 compounds enabled us to specify the level of confidence for each AC assignment and to get a more direct feeling for the quality of the assignment based on the position of the newly assigned molecule in previous successful assignments, which is a major advantage over the program suggested by Shen *et al.*

IV. Conclusion

In this paper a very fast and transparent method is suggested to quantify the confidence level of an assignment of absolute configuration. The method developed is based on user-independent, neighbourhood similarity measures for a quantum chemically computed and an experimental infrared spectrum. This similarity measure is extended to VCD by considering separately the similarity for the positive and the negative part of computed and experimental VCD spectra. This is done for both enantiomers and the absolute difference in VCD similarity with respect to experiment, is calculated. A successful assignment is characterized by high similarity between theory and experiment for one enantiomer and a low value for the other enantiomer.

In the next step, the procedure described above is applied for a large set of molecules.

It is found that for a large majority of molecules, the similarity measures immediately result in the correct enantiomer to have the highest similarity to experiment. Based on the results for the entire database, a numerical confidence level is computed that reflects the percentage of assignments made using the procedure that resulted in correct assignments.

In all cases, the algorithm presented should be used to answer the question of how reliable an assignment made, is. It cannot replace expert assignment but simply attaches a degree of confidence of the assignment and can possibly suggest routes for improvement of the experiment, calculation or assignment.

V. References

1. J. Gal, "Chiral drugs from a historical point of view", in *Chirality in Drug Research*, (Eds.: E. Francotte, W. Lidner), Wiley-VCH, Weinheim (Germany), **2006**, pp. 1-26.
2. F.J. Devlin, P.J. Stephens, J.R. Cheeseman, M.J. Frish, *J.Phys. Chem. A*, **1997,** *101*, 9912-9924.
3. L.A. Nafie, *Appl. Spectrosc.*, **1996,** *50*, 14A-26A.
4. P.J. Stephens, *J. Phys. Chem.*, **1985,** *89*, 748-752.
5. L.A. Nafie, T.B. Freedman, *J. Chem. Phys.*, **1983,** *78*, 7108-7116.
6. T. Kuppens, W. Herrebout, B. Van der Veken, P. Bultinck, *J. Phys. Chem. A*, *110*, **2006,** 10191-10200.
7. T. Buffeteau, D. Cavagnat, A. Bouchet, T. Brotin, *J. Phys. Chem. A.*, **2007,** *111*, 1045-1051.
8. C. Cappelli, S. Monti, A. Rizzo, *Int. J. Quant. Chem.*, **2005,** *104*, 744-757.
9. E. Debie, L. Jaspers, P. Bultinck, W. Herrebout, B. Van der Veken, *Chem. Phys. Lett.*, **2007,** *450*, 426-430.

10. E. Debie, P. Bultinck, W. Herrebout, B. Van der Veken, *Phys. Chem. Chem. Phys.*, **2008,** *10*, 3498-3508.
11. Stephens, P.J., "Vibrational Circular Dichroism: A new tool for the stereochemical characterization of chiral molecules", in *Computational Medicinal Chemistry for Drug Discovery*, (Eds. P. Bultinck, H. De Winter, W. Langenaeker, J.P. Tollenaere), Marcel Dekker, New York (USA), **2004**.
12. M. Rouhi, *Chem. Eng. News*, **2005,** *83 (29)*, 32.
13. T. Kuppens, K. Vandyck, J. Van der Eycken, W. Herrebout, B. van der Veken, P. Bultinck, *Spectrochim. Acta A: Molec. Bimolec. Spec.*, **2007,** *67*, 402-411.
14. T. Kuppens, W. Langenaeker, J. P. Tollenaere, P. Bultinck, *J. Phys. Chem. A*, **2003,** *107*, 542-553.
15. T. Kuppens, K. Vandyck, J. Van der Eycken, W. Herrebout, B.J. van der Veken, P. Bultinck, *J.Org. Chem.*, **2005,** *70*, 9103-9114.
16. T. Kuppens, PhD Thesis, Ghent University (Belgium), **2006**.
17. R. De Gelder, R. Wehrens, J.A. Hageman, *J. Comput. Chem.*, **2001,** *22*, 273-289.
18. J. Shen, C. Zhu, S. Reiling, R. Vaz, *Spectrochim. Acta Part A*, **2010,** *76*, 418-422.
19. I.M. Alecu, J.J. Zheng, Y. Zhao, D.G. Truhlar, *J. Chem. Theor. Comput.*, **2010,** *6*, 2872-2887.
20. P. Pulay, G. Fogarasi, G. Pongor, J.E. Boggs, A. Vargha, *J.Am. Chem. Soc.*, **1983,** *105*, 7037-7047.
21. C.N. Guo, R.D. Shah, R.K. Dukor, T.B. Freedman, X.L. Cao, L.A. Nafie, *Vibrat. Spec.*, **2006,** *42*, 254-272.
22. F.J. Devlin, P.J. Stephens, *J. Phys. Chem. A*, **1999,** *103*, 527-538.
23. F.J. Devlin, P.J. Stephens, *J. Am. Chem. Soc.*, **1999,** *121*, 7413-7414.
24. E. Debie, P. Bultinck, L.A. Nafie, R.K. Dukor, CompareVOA , 2010, BioTools, Inc.,

Jupiter, Florida (USA)

25. C. Cheng, G. Maggiora, M. Lajiness, M. Johnson, *J. Chem. Inf. Comp. Sc.*, **1996,** *36***,** 909.

FIGURES

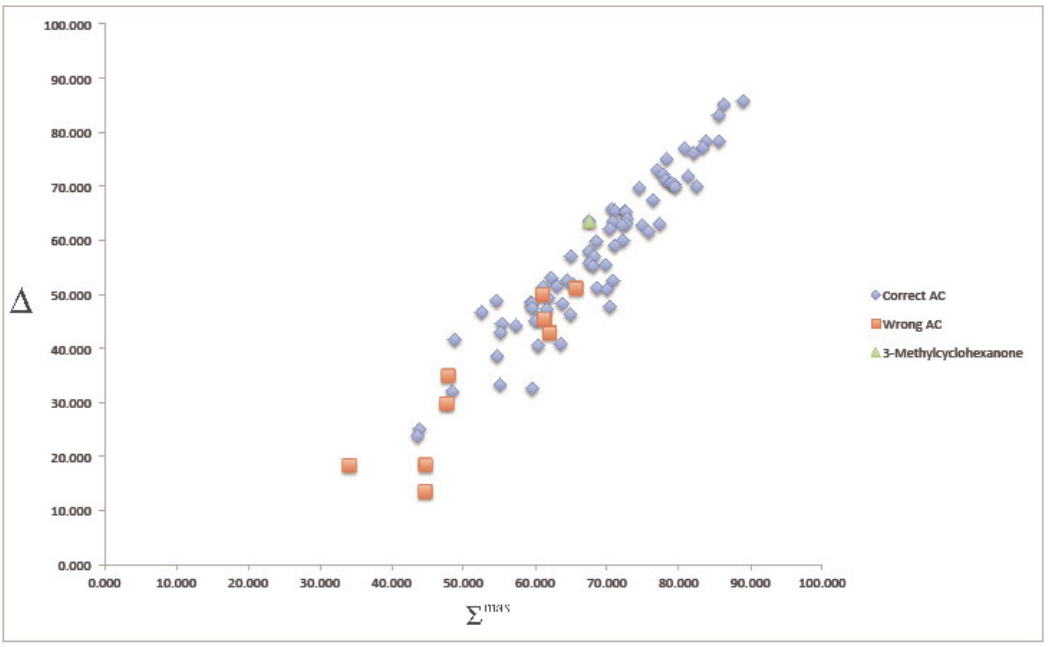


Figure 1: Plot of previous assignments in terms of Σ^{max} and Δ (both in %). The colour codes reflect the extent to which application of the current procedure led to the correct conclusion (black) or the wrong conclusion (red) on the AC with respect to more elaborate VCD assignments. The green marker indicates the position of a new AC assignment with respect to the database (see text).

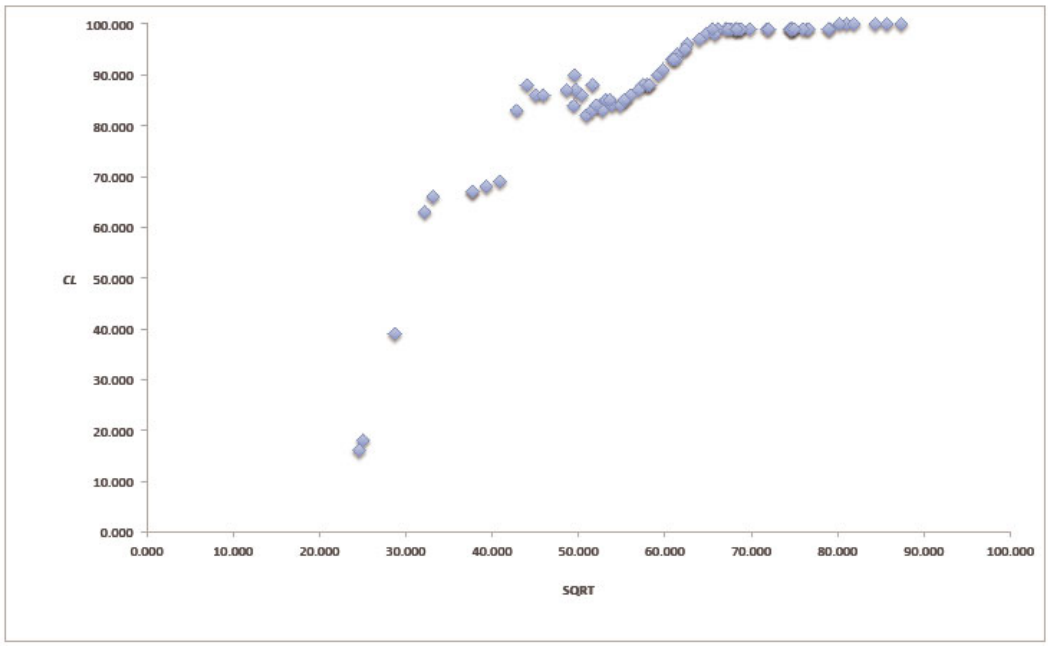
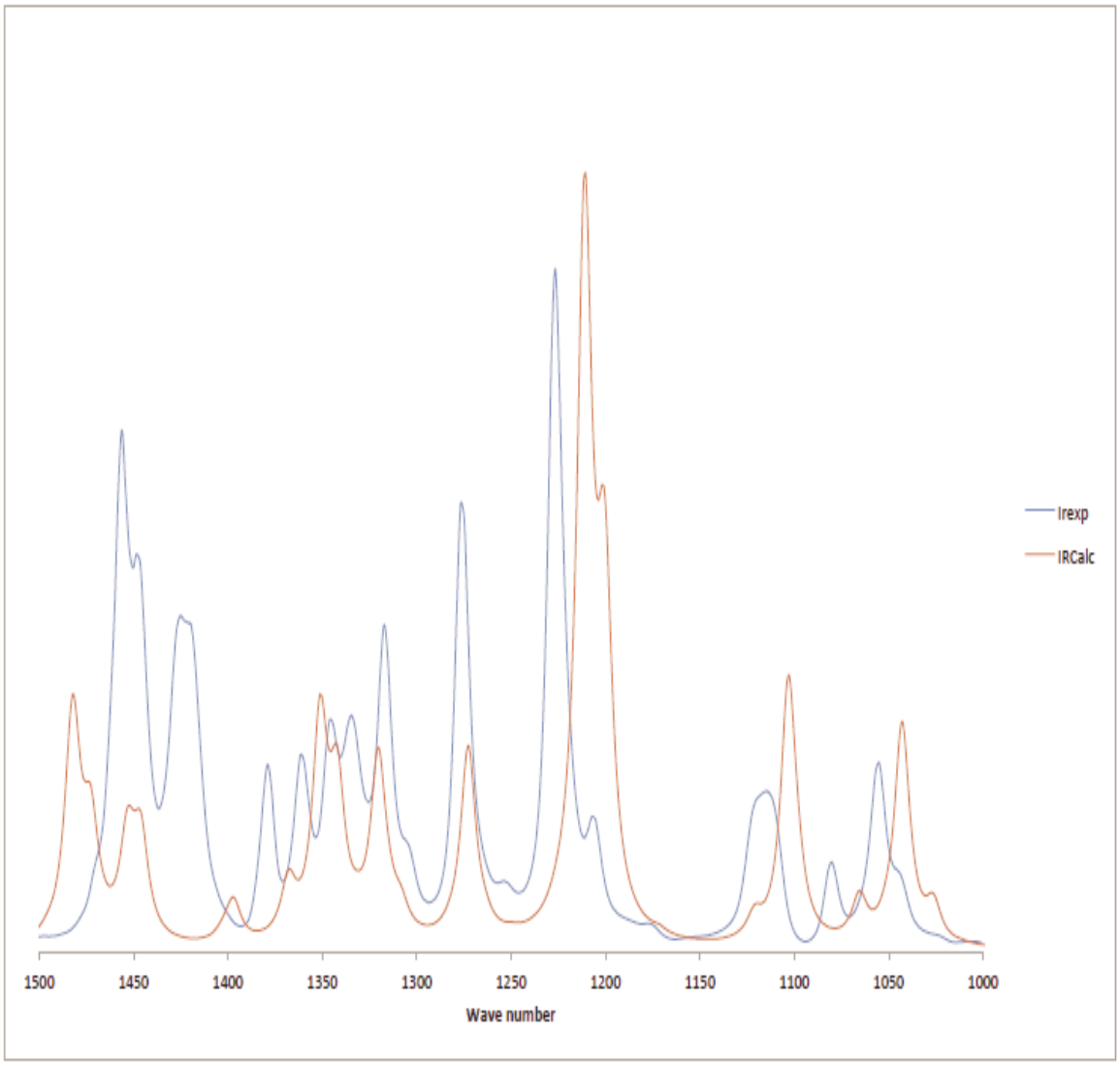
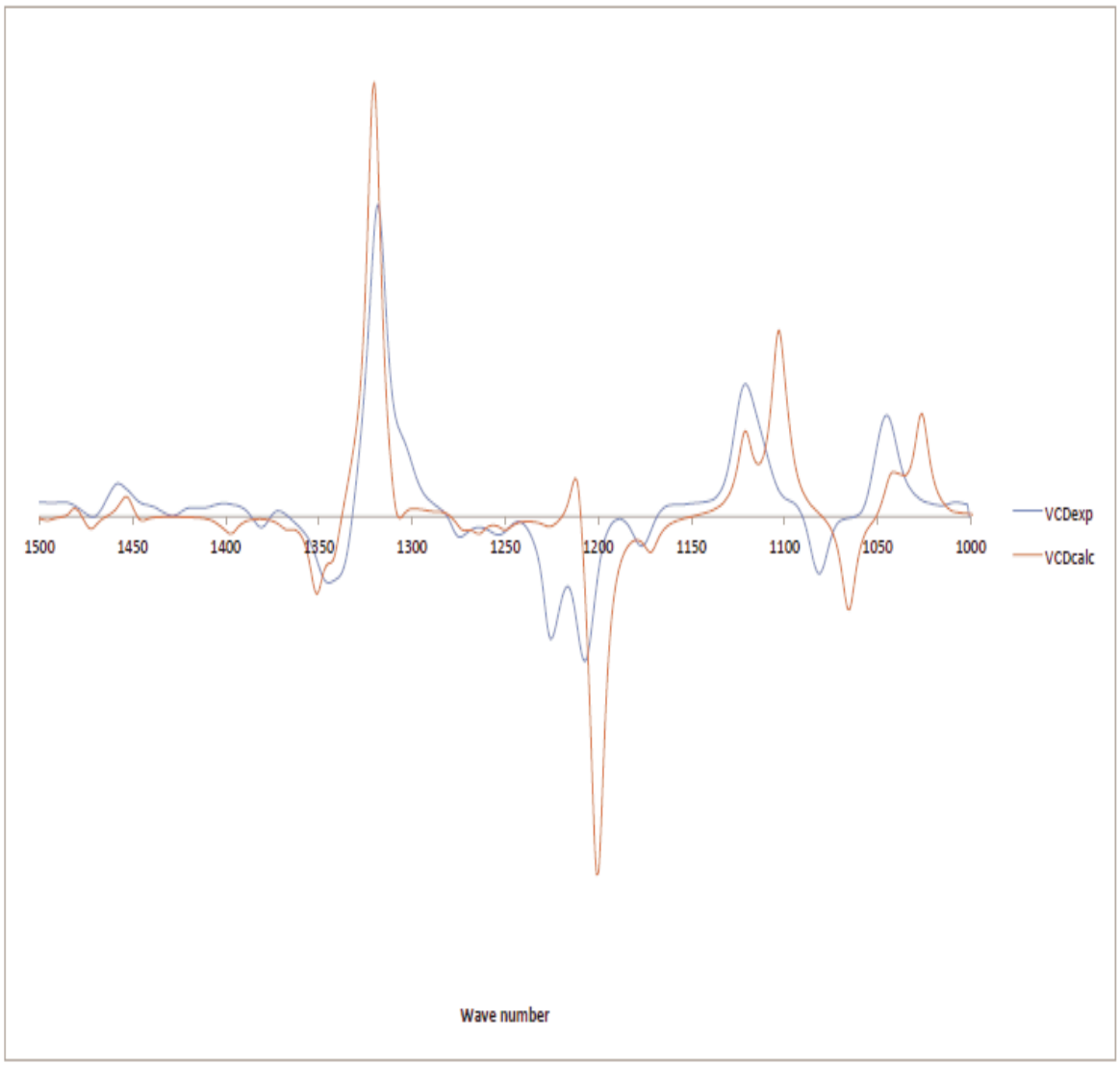


Figure 2: Variation of CL (in %) as a function of the generalized coordinate SQRT (see text, in %).


Irexp
IRCalc
1500
1450
1400
1350
1300
1250
1200
1150
1100
1050
1000
Wave number

(a)



(b)

Figure 3: Calculated and observed IR (a) and VCD (b) spectra for (+)-3R-methylcyclohexanone.

Table of contents text

Vibrational Circular Dichroism and Raman Optical Activity are powerful spectroscopic techniques to establish absolute configuration (AC). Their reliability and objectivity is further enhanced by plotting how good a new assignment of AC is compared to earlier successful AC assignments.